TEC-0050

AD-A283 427

US Army Corps of Engineers
Topographic Engineering Center

Robust Image Understanding - Techniques and Applications

Azriel Rosenfeld

University of Maryland
Computer Vision Laboratory
Center for Automation Research
College Park, MD 20742-3275

DTIC
ELECTE
AUG 16 1994
D

October 1993

94-25686
25pg

Approved for public release; distribution is unlimited.

Sponsored by:
Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22201-1714

Monitored by:
U.S. Army Corps of Engineers
Topographic Engineering Center
7701 Telegraph Road
Alexandria, Virginia 22310-3864

DTIC QUALITY INSPECTED 1

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
October 1993

3. REPORT TYPE AND DATES COVERED
Technical 1989 - 1992

4. TITLE AND SUBTITLE
Robust Image Understanding - Techniques and Applications

5. FUNDING NUMBERS
DACA76-89-C-0019

6. AUTHOR(S)
Azriel Rosenfeld

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Maryland
Computer Vision Laboratory
Center for Automation Research
College Park, MD 20742-3275

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Advanced Research Projects Agency
3701 North Fairfax Drive, Arlington, VA 22201-1714

U.S. Army Topographic Engineering Center
7701 Telegraph Rd., Alexandria, VA 22315-3864

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
TEC-0050

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The research for this report deals with the following topics: feature detection, including both edge detection and line slope estimation; robust and Bayesian estimation of image parameters; image and pattern matching; region segmentation and shape recognition; shape recovery and pose estimation; robot hand-eye coordination; robot navigation and path planning; visibility and navigation on terrain; motion transparency and motion parameter estimation; structure from motion, using both feature-based and normal-flow-based methods and regularization techniques.

14. SUBJECT TERMS
Edge detection, slope estimation, robust estimation, Bayesian estimation, matching, segmentation, shape recognition, recovery, pose estimation

15. NUMBER OF PAGES
36

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT
UNCLASSIFIED

20. LIMITATION OF ABSTRACT
UNLIMITED

Contents

1. Feature Detection **1**

1.1. Edge detection 1

1.2. Slope selection 1

2. Estimation **2**

2.1. Robust estimation 2

2.2. Bayesian estimation 2

3. Matching **3**

4. Segmentation and Recognition **7**

5. Recovery **8**

6. Hand-Eye Coordination **8**

7. Motion Planning **9**

8. Visibility and Navigation **10**

9. Motion Perception **11**

9.1. Transparency 11

9.2. Uncertainty and clustering 13

10. Structure from Motion **14**

10.1. Feature-based methods 14

10.2. Regularization methods 15

10.3. Normal flow based methods 16

11. Technical Reports **18**

Accesion For

NTIS CRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification .....

By .....
Distribution /

Availability Codes

Dist	Avail and / or Special
A-1	

PREFACE

This research is sponsored by the Advanced Research Projects Agency (ARPA), 3701 North Fairfax Drive, Arlington, Virginia 22201-1714 and monitored by the U.S. Army Topographic Engineering Center (TEC), Alexandria, Virginia 22315-3864, under Contract DACA76-89-C-0019, by the University of Maryland, College Park, MD 20742-3275. The ARPA point of contact is Dr. Oscar Firschein. The Contracting Officer's Representative at TEC is Ms. Lauretta Williams.

1. Feature Detection

1.1. Edge detection

A new robust algorithm for edge detection has been developed [22]. The algorithm detects both roof and step type edges. A pixel is declared as an edge pixel if there is a consensus between different processes that try to determine if the pixel lies on a discontinuity. A robust estimation method was used to estimate local fits to windows in the pixel's neighborhood and accumulate votes from each fit. The use of robust estimators makes it possible to transform any window possibly containing a discontinuity to a binary window containing a step edge in the location of the discontinuity. Conventional methods to detect this step edge can then be employed.

Experimental results were obtained on simulated edges and synthetic images with varying Gaussian and random noise levels, and the probability of detection was analyzed. The algorithm has also been applied to several real intensity and range images and has performed well. An example, including a comparison with the Canny edge detector, is given in Figure 1.

Figure 1: Comparison of the consensus-based (middle) and Canny (right) edge detectors applied to a noisy range image of a cube.

Another edge detection study [9] dealt with mask-based edge detectors. The orthogonal set of 3×3 Frei-Chen edge detection masks was originally proposed based on a vector space approach. The way the masks were chosen was not fully explained. An interpretation of the Frei-Chen masks has been formulated in terms of eight-dimensional Fourier transform coefficient vectors. The linear transformation between the nine-dimensional Frei-Chen space and the eight-dimensional Fourier transform space has been derived. A modified set of eight orthogonal masks based on the frequency space analysis was also developed.

1.2. Slope selection

A set of n distinct points in the plane defines $\binom{n}{2}$ lines by joining each pair of distinct points. The median slope of these $O(n^2)$ lines was proposed by Theil as a robust estimator for the slope of the line of best fit for the points. A randomized algorithm for selecting the kth smallest slope of such a set of lines which runs in expected $O(n \log n)$ time has been defined [10]. An efficient implementation of the algorithm was developed and used extensively to gain practical experience.

The problem of fitting a straight line to a set of data points is an important task in many application areas (e.g., statistical estimation, image processing, and pattern recognition). Recently the computation of linear estimators that are *robust* has been recognized as important, since these estimators are insensitive to outlying data points, which arise often in practice. One such robust estimator studied [42], the *repeated median* line estimator, achieves the highest possible breakdown point of 50%. The following results were obtained: (1) a simple practical randomized algorithm that runs in $O(n \log^2 n)$ time with high probability, and (2) a slightly more complex randomized algorithm which performs as well asymptotically, but empirical evidence shows that this algorithm performs in time $O(n \log n)$ on many realistic input distributions. Empirical evidence for the efficiency of this algorithm was obtained under a number of input distributions.

2. Estimation

2.1. Robust estimation

Data processing for scientific and industrial tasks often involves accurate extraction of theoretical model parameters from empirical data, and requires automated estimation methods that are robust in the presence of "noisy" (i.e., contaminated) data. *Robust estimation* is thus an important statistical tool that is frequently applied in numerous fields of science and engineering (e.g., automated manufacturing, robotic navigation, image processing, and computer vision).

Since the computational complexity of a robust estimator is one of the most important measures of its practicality, searching for methods that reduce the time (and space) complexity of robust estimators is a desirable research goal. Several *computationally efficient* algorithms were developed [43] for the *exact* computation of robust statistical estimators. In particular, the design and analysis of such algorithms were studied for various problem domains, including line, curve, and surface fitting.

A general underlying methodology was introduced for the efficient computation of the classes of estimators considered. Specifically, *computational geometry* techniques in the derivation of robust estimation algorithms were applied. Furthermore, it has been demonstrated that the derivation, in particular, of *randomized algorithms* for the above tasks results in algorithms that have the following properties: (1) they always terminate and return the *correct* computational results, (2) the improved (expected) running times occur with extremely high probability, (3) they are quite easy to implement; (4) constants of proportionality (hidden by the asymptotic notation) are small (i.e., the algorithms are *practical*), and (5) they are space optimal (i.e., they require linear storage).

Implementational issues were considered in great detail and have resulted in considerable practical experience with the algorithms.

2.2. Bayesian estimation

Bayesian estimation has many applications in computer vision. A frequent objection to Bayesian estimation is that the probability density functions (pdf's) involved are usually not

exactly known. In fact, however [20], exact knowledge of the pdf's is not important; it often suffices to know the pdf's approximately. Furthermore, it may even suffice to have a family of pdf's, one of which approximates the actual pdf, provided a "second-stage" pdf on the family is specified such that the approximation of the actual pdf has high probability.

Bayesian estimation of digital signals is ordinarily concerned with the problem of estimating an ideal signal, given a noisy signal. The problem of partial or "qualitative" Bayesian *description*, rather than complete estimation of the ideal signal, was investigated [31]. For example, in the case of a piecewise constant signal, instead of estimating the value of the ideal signal, one can seek only a piecewise symbolic description of the signal—e.g., is the value high or low, where these descriptors are defined by probability densities on the possible signal values. This task is computationally less costly than that of complete Bayesian estimation of the signal; moreover, it has been found that the descriptions can be estimated robustly. This approach has been illustrated both for digital signals and for a simple class of digital images.

The problem of estimation using partial (e.g., compressed) information about the observations is important in practice. One reason for its importance is that one might be interested in communicating data from the sensor(s) to the place where decisions are made (e.g., remote sensing data). Another reason is that estimation using compressed information might be less costly in terms of computation. The problem of estimating the parameters of a signal having known form was studied [35] (e.g. polynomial of degree r), using a Bayesian approach to estimation. In particular, conditions were studied under which the estimates obtained using partial information are the same as those obtained using full information. Also considered was an application to distribute detection (sensor fusion). The use of partial information to obtain partial estimates was also discussed.

3. Matching

Point-pattern matching relaxation techniques have been extended to allow matching of both point-like and linear features [17]. Specifically, a compatibility function was defined that relies on relative orientation information, which is translation and rotation invariant and can be more reliably extracted from noisy images than can positional information. This function was used to generalize the matching technique of Ranade and Rosenfeld; it can also be incorporated into other relaxation algorithms. The performance of the function has been illustrated using examples from the domain of object recognition in synthetic aperture radar (SAR) imagery. An example is shown in Figure 2.

Also developed was a computational vision approach [36] for the estimation of 2D translation, rotation, and scale from two partially overlapping images. The approach results in a fast and novel method that produces excellent results even when large rotation and scaling have occurred between the two frames, and the images are devoid of significant features. An illuminant direction estimation method is first used to obtain an initial estimate of camera rotation. A small number of feature points are then located based on a Gabor wavelet model for detecting local curvature discontinuities. An initial estimate of scale and translation is obtained by pairwise matching of the feature points detected in both frames. Finally, hierarchical feature matching is performed to obtain an accurate estimate of translation, rotation

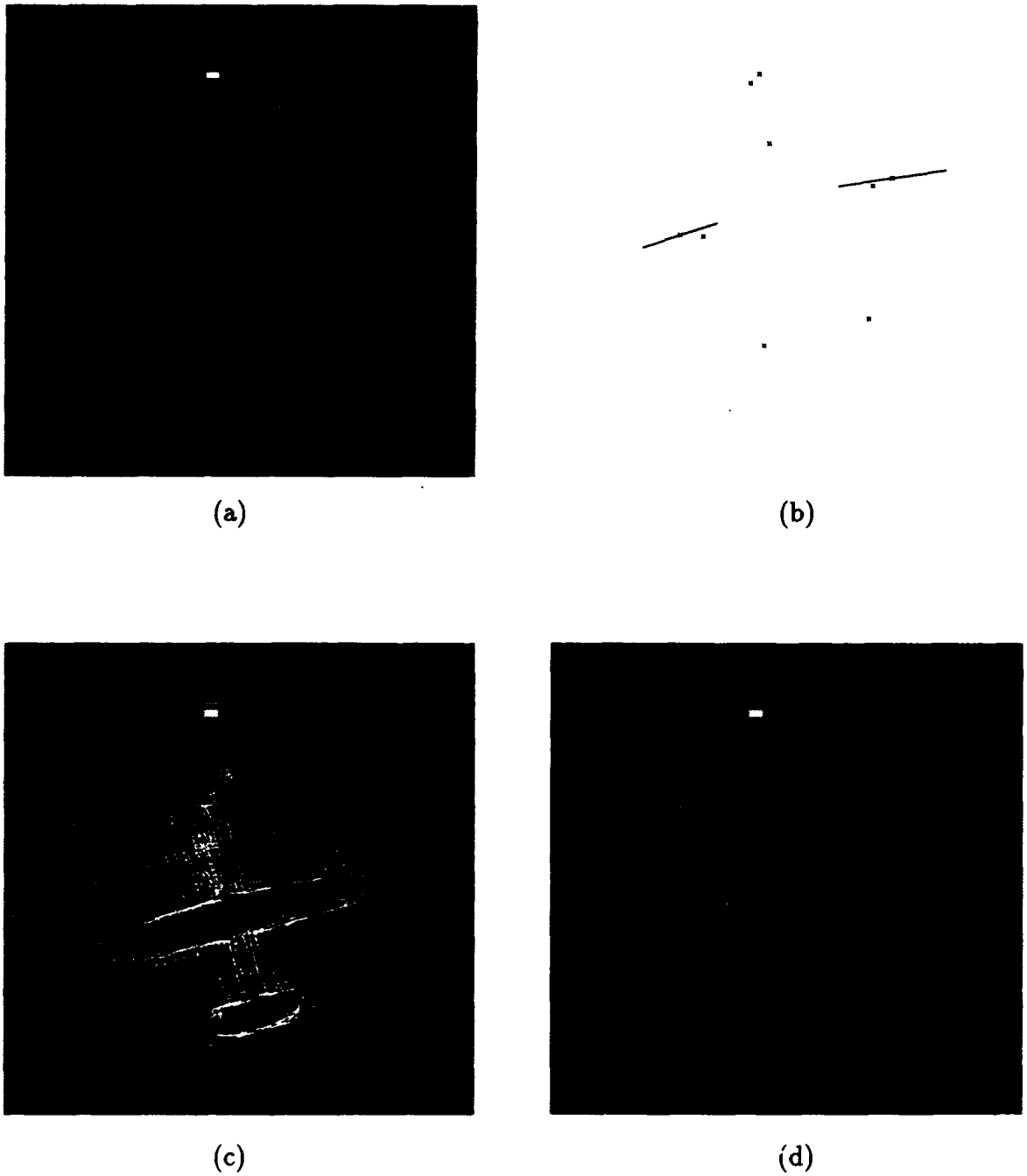


Figure 2: (a) Synthetic SAR image of a jet airplane. (b) Point and line features extracted from the image in (a). (c) Plausible configurations derived from high-confidence pairings after two iterations. (d) Plausible configurations after eight iterations.

and scale. Experiments with synthetic and real images have shown that this algorithm yields accurate results when the scales of the pair of images differ by up to 10%, the overlap between the two frames is as small as 35%, and the camera rotation between the two frames is significant. Experimental results on several real Mojave desert images acquired from a balloon have been obtained. The method has also been applied to texture and stereo image registration, satellite image mosaicking, and moving object detection. Two examples are shown in Figures 3 and 4.

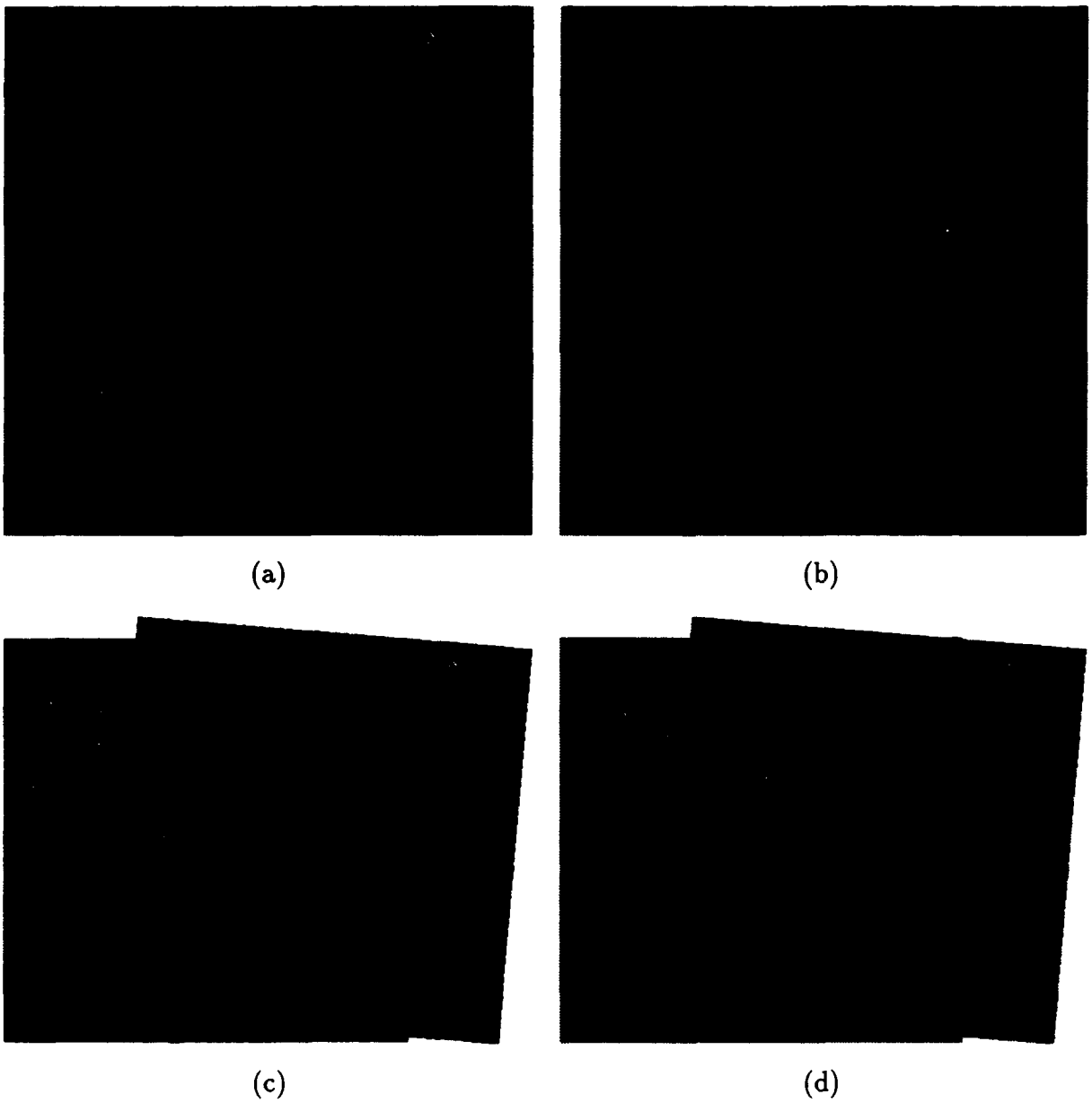
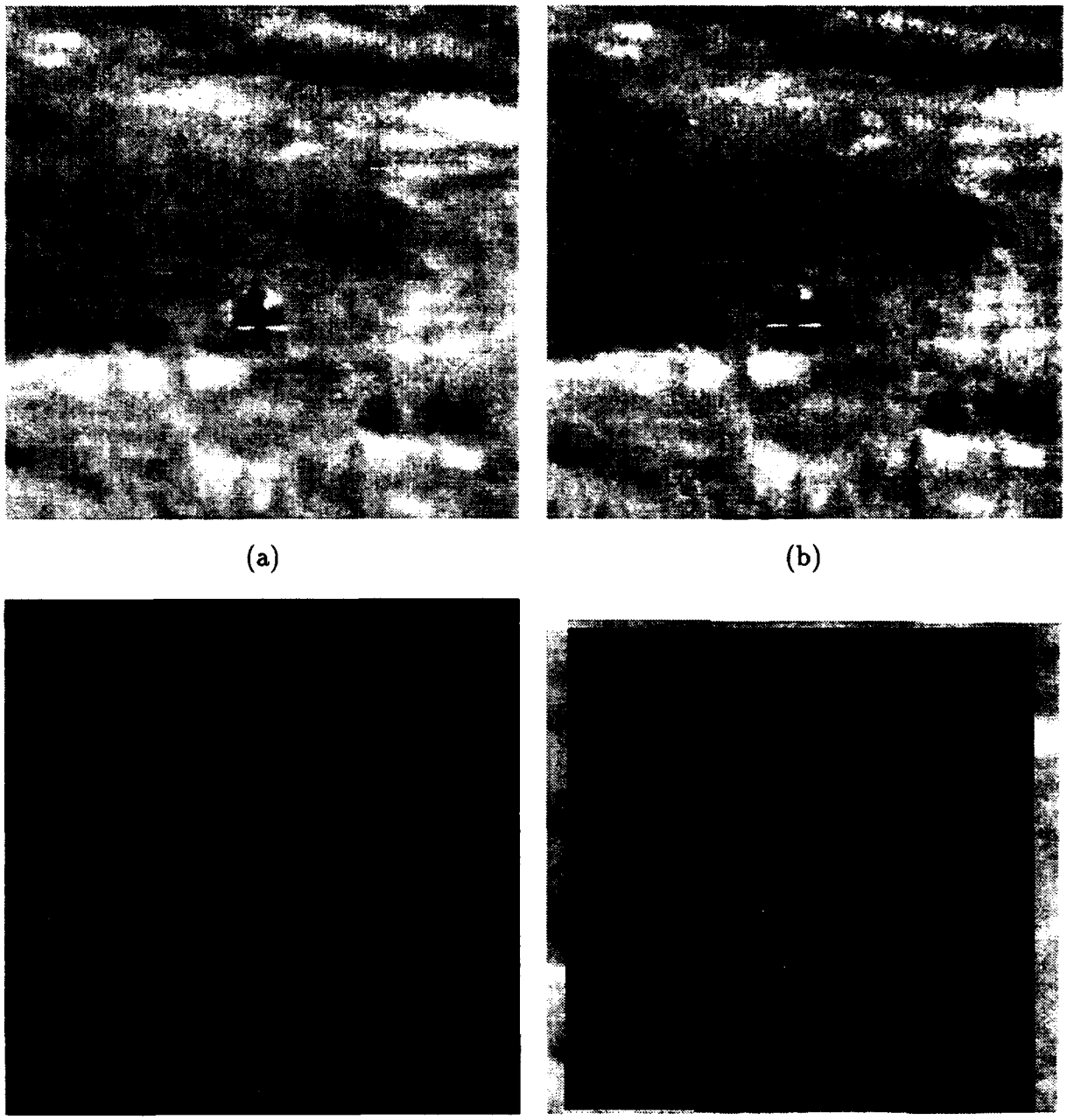


Figure 3: (a) & (b) Input images (Mojave desert). (c) Mosaicking of the two images. (d) Difference between the registered images.



(a) (b)

(c) (d)

Figure 4: (a) & (b) Two frames of a motion sequence. (c) Direct difference between (a) and (b). (d) Difference between the registered images.

4. Segmentation and Recognition

A method of recognizing compact objects in an image by energy function minimization was developed [3]. The energy function is based on a polar coordinate object representation, defined using any center from which the object's contour is visible. It incorporates both low-level and high-level information about the object: contour sharpness and smoothness at the low level, and contour shape at the high level. An example of the performance of the method is shown in Figure 5. Note how the center shifts to follow the centroid of the contour.

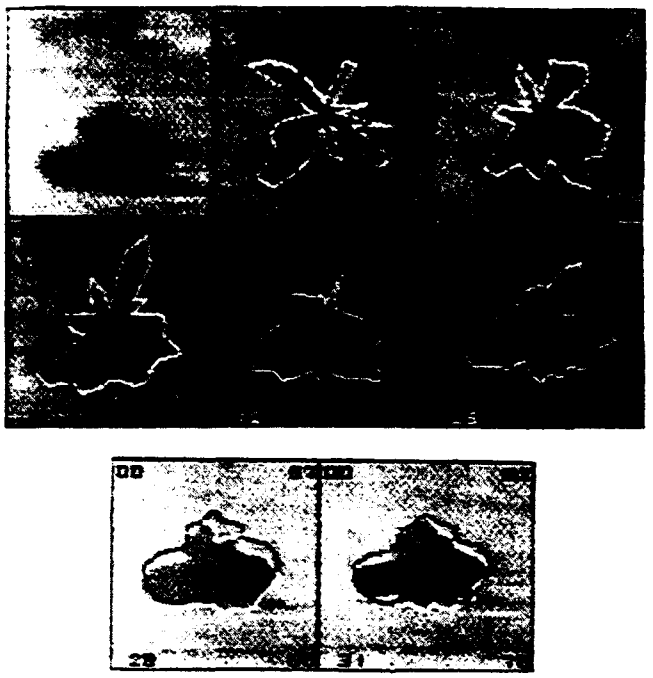


Figure 5: Example of object delineation and identification using simulated annealing. Upper left: Input image (tank in an infrared scene); black dot shows initial center. Successive frames show iterations 10, 20, . . . , 70 of the process; the white curve is the current estimate of the sharpest, smoothest contour, and the black curve is the best-fitting target model.

A shape recognition method was developed [7] based on an intrinsic equation representation of the 2D silhouette of a shape. This representation provides a method of recognition that is insensitive to perspective distortion and also allows the slant of the shape to be estimated. A parameter called the "tolerance" is incorporated in the method, which makes it possible to change the scale (relative resolution) of shape processing.

The presence of an object in an image usually does not depend on its position within the visual field. That is, its presence is *invariant* with respect to such properties as translation, rotation, and size. This presents problems for learning algorithms whose only feedback involves the existence of the target object, not its position. It must correctly determine an input-output behavior without knowing exactly which inputs are relevant to the behavior at any point in time. The *constraint motion* learning algorithm was applied to the problem of invariant learning [4]. The properties of the algorithm facilitate correct learning in distributed

environments and help with learning under invariance. A hierarchical learning scheme was formulated that improves accuracy without significantly increasing spatial requirements.

More recently, the problem of object recognition was studied [39] by considering it in the context of an agent operating in an environment, where the agent's intentions translate into a set of behaviors. In this context, an object can fulfill a function; if the agent recognizes this, it has in effect recognized the object. What is needed to perform this type of recognition is, on one hand, a definition of the desired function, and on the other, the means of determining whether the object can fulfill that function. To find out if an object can fulfill a function, it is necessary to perform various partial recovery tasks; in other words, it is only necessary to solve subproblems of the general recovery problem.

5. Recovery

Plants, such as trees, can be modeled by three-dimensional hierarchical branching structures. If these structures are sufficiently sparse, so that self-occlusion is relatively minor, their geometrical properties can be recovered from a single image. Specifically, it has been shown [40] that the parameters of a classical tree branching model can be recovered from a single orthographic image of a tree.

The pose of an object can be found from a single image when the relative geometry of four or more noncoplanar visible feature points is known. An algorithm was developed [41], called *POS* (Pose from Orthography and Scaling), that solves for the rotation matrix and the translation vector of the object. It uses a linear algebra technique under the scaled orthographic projection approximation. A second algorithm, *POSIT* (POS with ITerations), uses the pose found by POS to remove the perspective distortions from the image, and then applies POS to the corrected image instead of the original image. POSIT converges to accurate pose measurements after a few cycles of image corrections and POS computations, even in conditions where perspective distortions are large. POSIT can be used with many feature points at once for added insensitivity to measurement errors and image noise. POSIT can be implemented in 25 lines or less in Mathematica.

6. Hand-Eye Coordination

Traditional approaches to robot hand/eye coordination require that various components of the system be calibrated with respect to a common reference, but calibration is difficult and error-prone and may invalidate the complex, high-precision inverse kinematic computations that are also a feature of these approaches. A fundamentally new control technique was developed [16] that does not require any calibration and closely integrates visual feedback into the control mechanism. This is made possible by the introduction of a mapping, called the *Perceptual Kinematic Map*, from the control space of the manipulator directly onto a space defined by a set of *measurable* image parameters. This strategy achieves robustness by monitoring qualitative rather than quantitative changes as it explores the surface defined by this mapping. Furthermore, it employs a Kalman-Bucy filter for additional robustness in measuring image parameters. Successful experimental results were obtained, and possible generalizations and extensions of the technique were considered.

A general framework was developed [32] for reasoning about robot hand positioning tasks involving a moving target, such as catching, hitting, interception, etc. It has been shown how this framework may be used to achieve robust vision-based control. Different levels at which visual input is involved were considered in pursuing the dynamically-defined goal. A given task is first transformed into one of constrained trajectory planning on a topological space defined by a set of image parameters. A learning phase first learns the qualitative features of this *perceptual control surface* so that further operations may be carried out autonomously without precise calibration of different parts of the system. This differs significantly from the classical approaches that require more accurate descriptions of the robot environment and the manipulation task.

7. Motion Planning

Current approaches to robot motion planning are limited in their ability to deal with an uncertain and dynamically changing environment. Difficulties involved in modeling the situation were analyzed and a probabilistic model was developed based on *discrete events* that abstract the dynamic interaction between the mobile robot and the unknown part of the environment. The resulting framework makes it possible to design and evaluate motion planning strategies that consider both the known portion of the environment and the portion that is unknown, but satisfies a probability distribution. Three instances of the general model were studied [38] that yielded useful results in designing efficient motion planning algorithms as functions of parameters representing a robot's environment and its behavior with respect to unexpected events.

Specifically investigated [5] was the problem of robot navigation in the presence of moving obstacles and on the basis of visual information. A computational theory was developed that suggests several strategies that a robot can follow in order to plan a path (from a specified start to a specified end point) in the presence of moving obstacles, whose motion is not known a priori. The input to this perceptual process is time varying imagery acquired by the robot that navigates. The output is a strategy that indicates how the robot should move in order to obtain a safe path, i.e. a strategy that maximizes the probability of safely reaching the goal using visually acquired knowledge at every time instant. Smooth acceleration strategies for planning trajectories in 2D were also studied. Heuristics which approximate the minimax trajectory for a component of the acceleration have also been investigated.

In another study [13], the problem of efficiently planning a path for a robot between two points was addressed when the path is forced to change dynamically by the occurrence of certain *events* in the environment. An *event*, for example, may be the discovery of another moving object on a collision course with the robot. The robot is forced to take evasive action whenever such an *alarm* occurs. A probabilistic model was developed that represents the dynamic behavior in terms of *alarms* following a Poisson distribution, and *safety rules* that assume that some regions are *safe*. A provably optimal expected solution for the problem has been derived. The effect of the probabilistic parameter (λ) of the dynamic environment on the optimal path, and the use of "vision" (or *time to collision*) on the planned paths, have been studied. The results can be used in designing heuristics for path planning in a more general framework, and can be generalized to other situations. This study has given

insights into the role of various parameters on the *average* efficiency of path-planning in a simply dynamic, unknown environment. The simplicity of the model used is justified by the difficulty of analyzing a more complicated (unknown) dynamic environment, and by the generality of the results obtained using this simple model.

Finally, the problem of efficient path planning was studied [30] for a point robot in a partially known dynamic environment. The static known part of the environment consists of point shelters distributed in planar terrain, and the dynamic, unknown part is abstracted in the form of *alarms* that cause the robot to leave its current (pre-planned) path and divert to the nearest shelter. A probabilistic analysis was performed of the expected times for the dynamic paths generated when the alarms follow a Poisson distribution with parameter λ. A case study with three shelters was used to illustrate the dependence of the expected travel times on λ for two alternate static paths. Two different strategies were formulated for the general case of n shelters and shown to be superior for different ranges of values of the alarm rate λ (very low and very high values respectively). Some ways of generalizing the approach were also considered and possible applications have been examined.

In further studies [28, 29], a probabilistic method was developed for noisy sensor based robotic navigation in dynamic environments. The method generates an optimal trajectory by considering as optimal criteria, the probability of not colliding with the obstacles and the probability of accessing an operational position with respect to a moving target object. In particular, it can generate a trajectory that guarantees a tolerable associated collision risk. Estimates of the obstacle's kinematic parameters and measures of confidence in these estimates are used to produce the probability of collision associated with any robot displacement. The probability of collision is derived in two steps: a stochastic model is defined in the kinematic state space of the obstacles, and collision events are given simple geometric characterizations in this state space. In particular, the estimates can be used to define regions where the probability of encountering any obstacle is bounded by a predefined value.

8. Visibility and Navigation

In a study of 2D visibility, a parallel algorithm was developed [24] for computing the visible portion of a simple planar polygon with N vertices from a given point of the plane. The algorithm accomplishes this optimally for star-shaped polygons in $O(\log N)$ time using $O(N/\log N)$ processors. In the worst case, though, it may take $O(N \log N)$ time for oddly shaped polygons. The algorithm is rather simple compared to other visibility related algorithms, and has a very small run time constant, making the algorithm faster and more practical to implement than others. The inter-processor communication needed for this algorithm involves only local neighbor communication and scan operations (i.e. parallel prefix operations). Thus, the algorithm can not only be implemented on an EREW PRAM, but also on a hypercube connected parallel machine, which is a more practical machine model. The algorithm has been implemented on the Connection Machine, and various performance tests were conducted.

Representing natural terrain is an important issue in a variety of application domains. Various digital models have been developed that are able to represent terrain. Among them, regular grids have been extensively used because of their simplicity and because they can

be directly embedded on massive parallel architectures with fixed topologies. On the other hand, Triangulated Irregular Networks (TINs) better adapt to the irregular nature of natural terrain, but they do not offer any kind of regularity. A parallel algorithm was developed [26] to compute a TIN based on the Delaunay triangulation. The algorithm is designed for a massive SIMD computer with general communication, and has been implemented on a Connection Machine.

An algorithm was also developed [33] for solving region-to-region visibility problems on digital terrain models using massively parallel hypercube machines like the Connection Machine CM-2. This algorithm is an extension of an earlier developed point-to-region visibility algorithm. Since global communication is the bottleneck in this kind of algorithm, the algorithm focuses on the reduction of global communication. The algorithm analyzes a strip of the source region at a time, and sweeps through the source, strip by strip. At most, four sweeps are needed for the analysis. By exploring the coherence properties in the processor structure, global communication is minimized, and complexity is substantially improved. Furthermore, all global write operations are exclusive and concurrency in global read operations is minimized. Since the problem size is usually large, rules of decomposition have been designed to efficiently handle cases where the required number of processors is greater than available. The algorithm has been implemented on a Connection Machine CM-2, and results of computational experiments are presented.

On a more general level, a new type of visual information was formulated [34] which can be exploited by algorithms for path planning or obstacle avoidance. Traditionally, a robot's visual system is assigned the task of reconstructing the geometry of the surrounding scene. The navigation problem can then be solved by means of classical robotics (control of mechanisms). Unfortunately, it is still impossible to accurately compute the depth maps robots are supposed to use for navigating. Furthermore, it appears that such data may in fact not be the most suitable for the goals we want to achieve. A new approach to the navigation problem was developed, based on the exploitation of *free space doors*, in which visual processes are closely and actively integrated with the control of the robotic system.

Finally, an approach for autonomous localization of ground vehicles on natural terrain was developed [37]. The localization problem is solved using measurements including altitude, heading and distances to specific environmental points. The algorithm utilizes random acquisition of distance measurements to prune the possible location(s) of the viewer. The proposed approach is also applicable to airborne localization. The computational complexity of the implementation on the Connection Machine and the accuracy of the localization have been analyzed.

9. Motion Perception

9.1. Transparency

Two line patterns in relative motion [1] can give rise to either the perception of motion coherence or that of motion transparency. In the case of motion coherence, one velocity is perceived for both patterns, whereas for motion transparency, two velocities are perceived. The velocity histogram, which counts the number of occurrences of each observed value of the

velocity vector, is an important tool for the detection of coherence or transparency. When this histogram is unimodal, coherence is perceived, and when it is bimodal, transparency. This was demonstrated for various types of line patterns, composed of parallel or non-parallel line segments or of polygonal lines.

If two or more curve patterns are used in relative motion, the probability of the perception of motion transparency is high. This work [2] has led to an explanation of this phenomenon. The existence of regions of high curvature makes it possible to solve the aperture problem for each individual pattern. If the average curvature of the patterns is high, then the errors in the measurement of the normal velocity component and the curvature are proportionally low. If the velocity of each pattern is estimated through the velocity histogram, which counts the number of occurrences of each velocity value, then the pattern velocity will give rise to a distinct peak. The peak spread is proportional to the errors in the measurement of the normal velocity component and the curvature. For patterns with regions of high curvature, the peaks will exhibit small spreads, and therefore, different peaks will have small overlaps. The existence of distinct peaks in the velocity histogram gives rise to the perception of motion transparency. On the other hand, if the peaks have a large overlap, or if we have just one peak in the velocity histogram, then the perception of motion coherence results. In general, except for the case in which the average pattern curvature is very low, the different peaks have a small overlap, and therefore motion coherence is almost never perceived. This has been verified, through perceptual experiments, for different types of periodic open and closed curve patterns.

The task of segmenting multiple objects moving in space can require processing of the optical flow. This becomes especially difficult when small objects are densely distributed in space, like trees and bushes in a forest, or partially transparent objects. In this case motion transparency is perceived; this requires computing more than one value of the optical flow at each pixel, which is not accounted for by current motion theories. A statistical model has been developed [14] for the perception of motion transparency. The model has applied it to the analysis of situations involving two superimposed line patterns moving in the frontoparallel plane. If these patterns have regions of high curvature, or features like end-points or corners, the aperture problem can be solved for each pattern separately; consequently, motion transparency is perceived. On the other hand, in the absence of features, or for small curvature, motion coherence is perceived which is given by the motion of the compound pattern. A statistical model has been developed for the perception of motion transparency and coherence which is given by a two-stage process for the extraction of the optical flow and the velocity histogram. The velocity histogram, which is a plot of the number of occurrences of each velocity vector, is unimodal for motion coherence and bi-modal for motion transparency. The image is divided into regions, and inside each of them the optical flow is computed. The velocities of line end-points and corners are computed by matching them between images. For lines, the normal velocity components are combined by computing the intersection of the corresponding constraint lines in the velocity space. A generalized version of the two-stage process is used for the extraction of the optical flow which takes into account superimposed patterns. This model is also able to predict the transition between the perception of motion transparency and coherence, and it is in good agreement with informal perceptual experiments done with line patterns.

9.2. Uncertainty and clustering

Energy filters are tuned to space-time frequency orientations. In order to compute velocity it is necessary to use a collection of filters, each tuned to a different space-time frequency. In a probabilistic framework, the properties of the motion uncertainty have been analyzed [8]. Its lower bound, which can be explicitly computed through the Cramér-Rao inequality, will have different values depending on the filter parameters. It has been shown, for the Gabor filter, that in order to minimize the motion uncertainty, the spatial and temporal filter sizes cannot be arbitrarily chosen; they are only allowed to vary over a limited range of values. Consequently, the temporal filter bandwidth is larger than the spatial bandwidth. This property is shared by motion sensitive cells in the primary visual cortex of the cat, which are known to be direction selective and are tuned to space-time frequency orientations. It seems that these cells have larger temporal bandwidths relative to their spatial bandwidth because they compute velocity with maximum efficiency, that is, with minimum motion uncertainty.

Image motion can be estimated by matching feature "interest" points in different frames of video image sequences. The matching is based on local similarity of the displacement vectors. Clustering in the displacement vector space can be used [18] to determine the set of plausible match vectors. Subsequently, a similarity based algorithm performs the actual matching. The feature points are computed using a multiple filter image decomposition operator. The algorithm has been tested on synthetic as well as real video images. The novelty of this approach consists of the fact that it handles multiple motions and performs motion segmentation.

A method was developed [23] for the discrimination of 3D texture patterns through the use of motion cues. 3D texture is defined by the 3D distribution of primitive elements, or volumetric *texels*, which can be solid or planar, opaque or transparent. Trees and bushes are examples of 3D textures which are very common in natural scenes. One of the motivations to work in the domain of 3D texture comes from the fact that current theories of low-level vision, including theories of motion, stereo, and texture, are unable to deal with this kind of visual information. 3D texture patterns can be discriminated in time-varying imagery by using velocity information observed from their projections onto the image plane. The method of doing so combines the velocity information given by contours and features, such as end-points and corners. The image is divided into regions, and each region into windows. For each window the feature velocity is computed through correspondence; the normal velocity components are also measured and the intersection of all possible constraint lines is computed. The feature and contour velocities are used to generate the velocity histogram, which is the plot of the number of occurrences of each velocity vector. If a region contains two superimposed patterns in relative motion then its velocity histogram is bi-modal. This corresponds to the perception of motion transparency. This method has been successfully tested for both synthetic 3D textures and real images of plants. An example is shown in Figure 6.

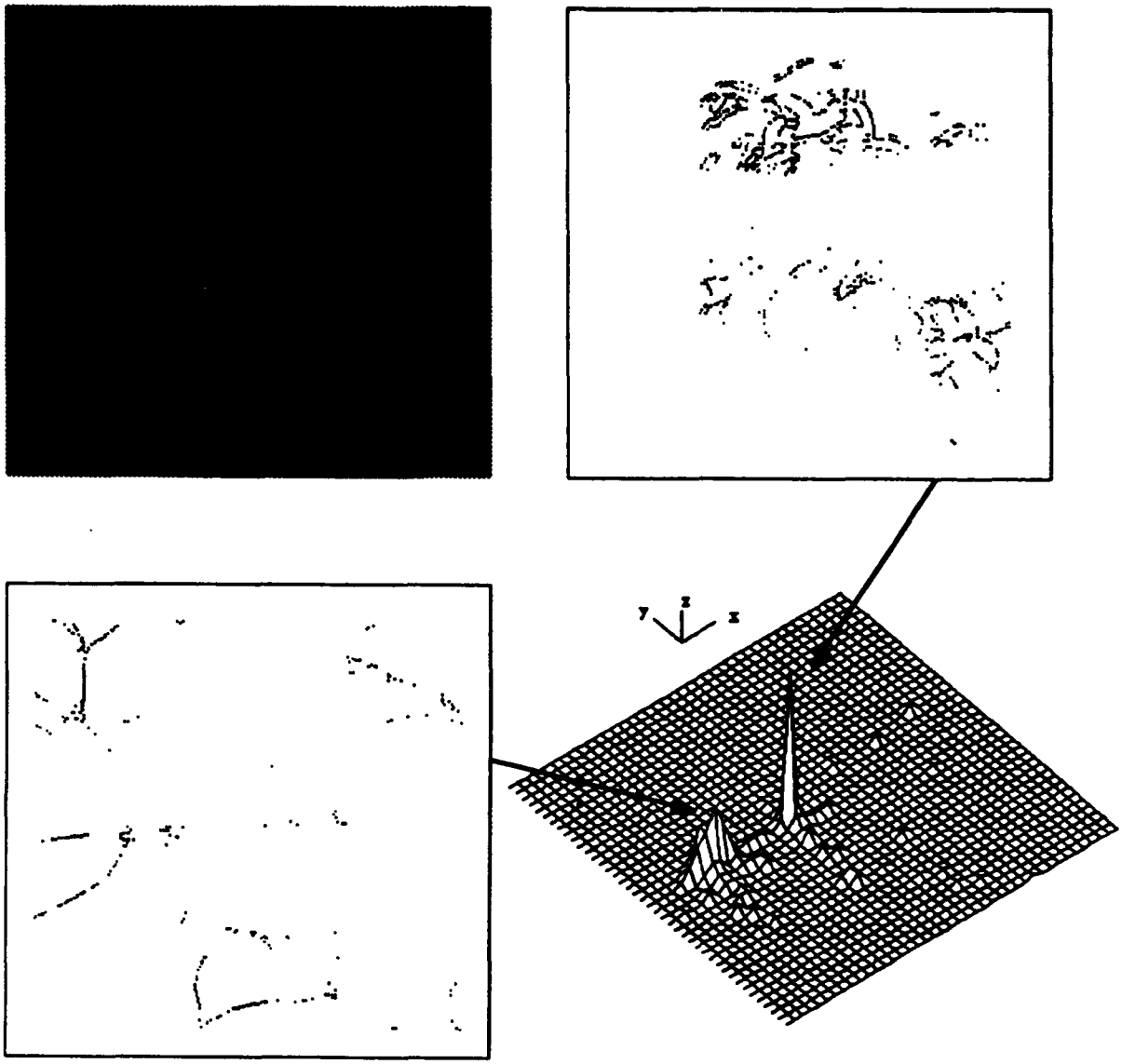


Figure 6: Upper left: Two bushes in front of one another. Lower right: Bimodal velocity histogram. Lower left: Edges contributing to the first peak, which belong to the closer bush. Upper right: edges contributing to the second peak, which belong to the farther bush.

10. Structure from Motion

10.1. Feature-based methods

The long sought linear algorithm was formulated [6] for the point and line correspondence problem. A new statistical definition of feature points was also introduced, under which point features and line features are just the two extremes of a spectrum of possible features. Almost any pixel in the image can be classified and used as a feature point in this scheme. Based on this definition, an optimal algorithm was designed for the structure from motion problem that can utilize information from across the whole image. The input to the algorithm is the image displacement, and its uncertainty at each pixel for a set of three frames. The only

assumptions used are rigidity and Gaussian noise in the image displacements. The outputs are the parameters of the motion between the frames and the structure of the scene.

The theory behind this approach is simple, can be extended in several ways (e.g. to multiple frames), and has been developed with noise stability in mind. However, more important is that the new statistical definition of the features relaxes the requirements for the image displacement computation. If the tangential component of a displacement cannot be computed, its uncertainty is set to infinity. The algorithm can tolerate infinite uncertainty for all the tangential components. In this way the aperture problem is avoided.

Two important structure from motion problems in recent years have been the point based and the line based problem (using image motion of points or lines to find 3D motion and structure). A considerable advance came from the development of linear algorithms for lines and points separately. However the solutions to these two problems could not be combined into a linear algorithm that uses points and lines together. Such an algorithm has now been developed [15]. This algorithm needs three frames and a combination of point and line correspondences that give enough constraints to solve the problem. Using redundant points and lines, the algorithm exhibits stability in the presence of noise. It has been tested with simulated data under a wide variety of conditions.

10.2. Regularization methods

Humans use various cues in order to understand the structure of the world from images. One such cue is the contours of an object formed by occlusion or from surface discontinuities. It is known that contours in the image of an object provide various amounts of information about the shape of the object in view, depending on assumptions that the observer makes. Another powerful cue is motion. The ability of the human visual system to discern structure from a motion stimulus is well known and it has a solid theoretical and experimental foundation. But when humans interpret a visual scene, they use various cues in order to understand what they observe, and the interpretation comes from combining the information acquired from the various modules devoted to specific cues. In such an integration of modules it seems that each cue carries a different weight and importance.

Several experiments were performed [11] in which the only cues available to the observer were contour and motion. It turns out that when humans combine information from contour and motion to reconstruct the shape of an object in view, if the results of the two modules—shape from contour and structure from motion—are inconsistent, they totally discard one of the cues and an illusion is experienced. Examples of such illusions have been constructed and the conditions have been identified under which they occur. Finally, a computational theory has been introduced for combining contour and motion using the theory of regularization. The theory explains such illusions and predicts many more. The same computational theory, when applied to retinal motion estimation, explains the effect of boundaries on the perception of motion that gives rise to a set of well known illusions described by Wallach.

Inverse problems in low-level vision tend to be ill-posed and smoothness assumptions (regularization) need to be made in order to obtain unique solutions that vary continuously as a function of the data. But the solution must not smooth over discontinuities in the image and it is necessary to take into account the fact that the probability distributions

of the smoothness measures are not known. The most popular theories of discontinuous regularization (Blake, Marroquin) make strong assumptions about these distributions and also result in nonconvex optimization problems whose solutions are difficult to obtain or interpret. The theory of robust statistics (M-statistics) of Huber was applied [12] to obtain a convex regularization that is also maximally robust against misspecification of the probability distribution of large jumps in the unknown. This theory has been applied to the optical flow constraint, which is notoriously noisy and inaccurate. The results show that this convex regularization accurately preserves depth boundary information.

10.3. Normal flow based methods

An active observer can compute the relative depth of (stationary or moving) objects in the field of view using only the spatiotemporal derivatives of the time varying image intensity function. This can be done in a manner which is:

- purposive in the sense that it solves only the relative depth from motion problem and cannot be used for other problems related to motion; and
- active in the sense that the activity of the observer is essential for the solution of the problem. In fact, most of the computational burden is placed on the activity of the observer.

Results indicate [19] that exact computation of retinal motion (optic flow or displacements) does not appear to be a necessary first step for some problems related to visual motion, contrary to the conventional wisdom. In addition, it has been demonstrated that optic flow, whose computation is an ill-posed problem, is related to the motion of the scene only under very restrictive assumptions. As a result, the use of optic flow in some quantitative motion analysis studies is questionable.

Passive navigation refers to the ability of an organism or a robot that moves in its environment to determine its own motion precisely on the basis of some perceptual input, for the purposes of kinetic stabilization. A robust solution to the passive navigation problem was developed [25] which is purposive, in the sense that it does not claim any generality; it just solves the kinetic stabilization problem and cannot be used as it is for other problems related to 3D motion. The solution is qualitative, in the sense that it comes as the answer to a series of simple yes/no questions and not as the result of complicated numerical processing. Finally, it is active, in the sense that the activity of the observer (in this case "saccades") is essential for the solution of the problem.

The input to the perceptual process of kinetic stabilization that has been developed is the normal flow, i.e. the projection of the optic flow along the direction of the image gradient. Contributions of this work are the fact that translation can be estimated reliably from a normal flow field that also contains rotation, and the theoretical error analysis, which gives the method the potential of being used in a successful practical vision system.

When an object is moving in an unrestricted manner (translation and rotation) in the 3D world, in many cases, only the motion's translational components are of interest. For a monocular observer, using only the normal flow—the spatiotemporal derivatives of the

image intensity function—the problem of computing the direction of translation [27] has been solved. Optical flow is not used, since its computation is an ill-posed problem and it is not the same as the motion field—the projection of the 3D motion on the image plane—in the general case. Two methods have been developed that perform different operations on the normal flow; each of them requires the observer to be active. Both techniques address the problem in two consecutive steps. First, the direction of translation parallel to the image plane is determined, and it is then used to derive information about the motion in the third dimension. The activities which the observer must perform to solve this special problem are fixation and tracking: fixation, in order to simplify the reconstruction of 3D motion parameters for a small area in the image; and tracking, in order to compensate for the lack of existence of an optical flow field, and as a tool for accumulating 3D motion information over time.

11. Technical Reports

1. Azriel Rosenfeld, Radu S. Jasinschi and Helder J. Araujo, "Motion Transparency I: Straight Line Patterns." CAR-TR-465, CS-TR-2325, September 1989.

2. Radu S. Jasinschi, Azriel Rosenfeld and Helder J. Araujo, "Motion Transparency II: Curved Line Patterns." CAR-TR-477, CS-TR-2368, December 1989.

3. Noah S. Friedland and Azriel Rosenfeld, "Compact Object Recognition Using Energy Function Based Optimization." CAR-TR-478, CS-TR-2369, December 1989.

4. John R. Sullins, "Distributed Learning under Invariance." CAR-TR-479, CS-TR-2370, December 1989.

5. Anup Basu, "A Framework for Motion Planning in the Presence of Moving Obstacles." CAR-TR-481, CS-TR-2378, December 1989.

6. Minas Spetsakis and John (Yiannis) Aloimonos, "Unification Theory of Structure from Motion." CAR-TR-482, CS-TR-2379, December 1989.

7. Zygmunt Pizlo and Azriel Rosenfeld, "Recognition of 2D Shape in 3D Space." CAR-TR-484, CS-TR-2381, January 1990.

8. Radu S. Jasinschi, "Energy Filters, Motion Uncertainty, and Motion Sensitive Cells in the Visual Cortex: a Mathematical Analysis." CAR-TR-487, CS-TR-2397, February 1990.

9. Rae-Hong Park, "A Fourier Interpretation of the Frei-Chen Edge Masks." CAR-TR-489, CS-TR-2414, February 1990.

10. Michael B. Dillencourt, David M. Mount and Nathan S. Netanyahu, "A Randomized Algorithm for Slope Selection." CAR-TR-493, CS-TR-2431, March 1990.

11. John (Yiannis) Aloimonos and Liuqing Huang, "Motion-Boundary Illusions and their Regularization." CAR-TR-495, CS-TR-2434, March 1990.

12. David Shulman and Jean-Yves Hervé, "On the Regularization of Discontinuous Flow Fields." CAR-TR-498, CS-TR-2440, March 1990.

13. Rajeev Sharma, "Locally Efficient Path Planning in a Dynamic Environment with a Probabilistic Model." CAR-TR-507, CS-TR-2488, June 1990.

14. Radu Jasinschi, Azriel Rosenfeld and Kazuhiko Sumi, "The Perception of Visual Motion Coherence and Transparency: a Statistical Model." CAR-TR-512, CS-TR-2524, August 1990.

15. Minas Spetsakis, "Linear Algorithm for the Point and Line Correspondence Problem." CAR-TR-514, CS-TR-2528, September 1990.

16. Jean-Yves Hervé, Rajeev Sharma and Peter Cucka, "Qualitative Coordination of a Robot Hand/Eye System." CAR-TR-516, CS-TR-2544, October 1990.

17. Peter Cucka and Azriel Rosenfeld, "Linear Feature Compatibility for Pattern-Matching Relaxation." CAR-TR-530, CS-TR-2579, January 1991.

18. Amit Bandopadhay and John (Yiannis) Aloimonos, "Image Motion Estimation by Clustering." CAR-TR-533, CS-TR-2588, January 1991.

19. Liuqing Huang and (John) Yiannis Aloimonos, "Relative Depth from Motion Using Normal Flow: An Active and Purposive Solution." CAR-TR-535, CS-TR-2612, February 1991.

20. Mohamed Abdel-Mottaleb and Azriel Rosenfeld, "Inexact Bayesian Estimation." CAR-TR-542, CS-TR-2630, March 1991.

21. Saibal Banerjee, "An Optimal Algorithm to find the Degrees of Connectedness in an Undirected Edge-weighted Graph." CAR-TR-544, CS-TR-2638, March 1991.

22. Doron Mintz, "Robust Consensus Based Edge Detection." CAR-TR-546, CS-TR-2643, March 1991.

23. Radu S. Jasinschi, Azriel Rosenfeld, Peter Cucka and Kazuhiko Sumi, "Discriminating 3D Texture Patterns: The Velocity Histogram Method." CAR-TR-547, CS-TR-2644, March 1991.

24. Ling Tony Chen and Larry S. Davis, "A Parallel Algorithm for the Visibility of a Simple Polygon Using Scan Operations." CAR-TR-559, CS-TR-2691, June 1991.

25. Zoran Durić and (John) Yiannis Aloimonos, "Passive Navigation: An Active and Purposive Solution." CAR-TR-560, CS-TR-2692, June 1991.

26. Enrico Puppo, Larry Davis, Daniel DeMenthon and Yang-pin Ansel Teng, "Parallel Terrain Triangulation Using the Connection Machine." CAR-TR-561, CS-TR-2693, June 1991.

27. Cornelia Fermüller and (John) Yiannis Aloimonos, "Estimating 3D Motion from Image Gradients." CAR-TR-564, CS-TR-2706, June 1991.

28. Philippe Burlina, Daniel DeMenthon and Larry S. Davis, "Navigation with Uncertainty: I. Reaching a Goal in a High Collision Risk Region." CAR-TR-565, CS-TR-2717, June 1991.

29. Philippe Burlina, Daniel DeMenthon and Larry S. Davis, "Navigation with Uncertainty: II. Avoiding High Collision Risk Regions." CAR-TR-568, CS-TR-2724, July 1991.

30. Rajeev Sharma, David M. Mount and (John) Yiannis Aloimonos, "Probabilistic Analysis of Some Navigation Strategies in a Dynamic Environment." CAR-TR-569, CS-TR-2726, July 1991.

31. Mohamed Abdel-Mottaleb and Azriel Rosenfeld, "'Qualitative' Bayesian Estimation of Digital Signals and Images." CAR-TR-571, CS-TR-2731, August 1991.

32. Rajeev Sharma, Jean-Yves Hervé and Peter Cucka, "Analysis of Dynamic Hand Positioning Tasks Using Visual Feedback." CAR-TR-574, CS-TR-2734, August 1991.

33. Yang-pin Ansel Teng, Daniel DeMenthon and Larry S. Davis, "Region-to-Region Visibility Analysis Using Massively Parallel Hypercube Machines." CAR-TR-578, CS-TR-2754, September 1991.

34. Jean-Yves Hervé, "Dynamic Interpretation of Free Space Doors for Visual Navigation." CAR-TR-581, CS-TR-2758, September 1991.

35. Mohamed Abdel-Mottaleb and Azriel Rosenfeld, "Bayesian Estimation Based on Partial Information." CAR-TR-582, CS-TR-2759, September 1991.

36. Qinfen Zheng and Rama Chellappa, "A Computational Vision Approach to Image Registration." CAR-TR-583, CS-TR-2760, September 1991.

37. Yaser Yacoob and Larry S. Davis, "Computational Ground and Airborne Localization Over Rough Terrain." CAR-TR-591, CS-TR-2788, November 1991.

38. Rajeev Sharma, "A Probabilistic Framework for Dynamic Motion Planning in Partially Known Environments." CAR-TR-592, CS-TR-2791, October 1991.

39. Ehud Rivlin, (John) Yiannis Aloimonos and Azriel Rosenfeld, "Purposive Recognition: A Framework." CAR-TR-597, CS-TR-2811, December 1991.

40. Adlai Waksman and Azriel Rosenfeld, "Sparse, Opaque Three-Dimensional Texture, 1: Arborescent Patterns." CAR-TR-598, CS-TR-2812, December 1991.

41. Daniel F. DeMenthon and Larry S. Davis, "Model-Based Object Pose in 25 Lines of Code." CAR-TR-599, CS-TR-2815, December 1991.

42. David M. Mount and Nathan S. Netanyahu, "Computationally Efficient Algorithms for a Highly Robust Line Estimator." CAR-TR-600, CS-TR-2816, December 1991.

43. Nathan S. Netanyahu, "Computationally Efficient Algorithms for Robust Estimation." UMIACS-TR-92-54, CS-TR-2898, May 1992.